

Metoda wykrywania fałszywych alarmów w systemie monitorującym sieć telekomunikacyjną kopalni

Method for detection of false alarms in telecommunication network monitoring system of the mine



Prof. zw. dr hab. Jerzy Korczak^{*)}



Mgr Bartłomiej Karaban^{**)}

Treść: Zapewnienie bezpieczeństwa pracy i utrzymanie ciągłości wydobywania to kluczowe zadania systemów telekomunikacyjnych w górnictwie głębinowym. Systemy te pomimo nowoczesnych i innowacyjnych rozwiązań monitorowania infrastruktury nie są wolne od wad. Praktycznym problemem jest występowanie fałszywych alarmów o uszkodzeniu infrastruktury, które powodują negatywne skutki, takie jak zwiększenie kosztów bieżącej eksploatacji systemu i przeciążenie informacyjne operatorów. W publikacji zaproponowano metodę wykrywania fałszywych alarmów w systemie telekomunikacyjnym kopalni oraz zaprezentowano niektóre reguły dostarczające użytecznej wiedzy z danych. Eksperymenty zostały przeprowadzone na rzeczywistych danych pochodzących z systemu telekomunikacyjnego funkcjonującego w kopalni KGHM Polska Miedź S.A.

Abstract: The key task of telecommunication systems in deep mining is to ensure safety and continuity of production. These systems, despite modern and innovative infrastructure monitoring solutions, are not free from drawbacks. The practical problem is the occurrence of false alarms of damage to infrastructure, which cause many negative effects such as the increase in the cost of the current operation of the system, information overload of operators or service errors. This paper proposes a method for detecting false alarms in the communication system of the copper mine and presents some rules that provide useful knowledge extracted from the database. The experiments were carried out on real data from the telecommunication system operating in the mine KGHM Polska Miedz S.A.

Słowa kluczowe:

eksploracja danych, indukcyjne drzewa decyzyjne, klasyfikacja, system telekomunikacyjny

Key words:

data exploration, inductive decision trees, classification, telecommunication system

1. Wprowadzenie

Bezpieczeństwo pracy w kopalni, dobra organizacja oraz utrzymanie ciągłości produkcji wymaga sprawnego i efektywnego systemu monitorowania stanu instalacji telekomunikacyjnej, maszyn, urządzeń i pracowników. Jedną z istotnych funkcji systemu monitorującego sieć telekomunikacyjną jest gromadzenie i przekazanie operatorom informacji o statusie komunikacji z urządzeniami sieciowymi, o obecności napięcia zasilania, czasie i lokalizacji wystąpienia awarii i wartościach parametrów pracy urządzeń, oraz o zagrożeniach zaistnienia awarii [2]. Informacja o awarii dla operatora systemu oznacza

utrata łączności w wyrobiskach, w którym te urządzenia są zainstalowane.

Przedstawione w artykule badania dotyczyły znanego i trudnego problemu występującego w automatycznych systemach monitorujących – problemu identyfikacji fałszywych alarmów, braku komunikacji z wybranym urządzeniem infrastruktury sieciowej lub ich grupą [5]. Zaznaczmy, że systemy monitorujące w kopalni, i to potwierdzają też nasze badania, zawsze generują alarm o braku łączności w przypadkach rzeczywistej jej utraty [4].

Rozpoznanie fałszywych alarmów i minimalizacja ich wystąpień jest jednym z aktualnych problemów projektowania i funkcjonowania automatycznych systemów monitorujących. Jednym z kryterium kwalifikacji alarmu jest jego czas trwania. W praktyce w przypadku krótkiego czasu trwania alarmu, operatorzy traktują alarm jako fałszywy. Takie podejście skutkuje

^{*)} Uniwersytet Ekonomiczny we Wrocławiu ^{**)} Uniwersytet Ekonomiczny we Wrocławiu, KGHM Polska Miedź S.A.

„wyczekiwaniem” na zakończenie stanu alarmowego (przywrocenie komunikacji), co powoduje opóźnienia w zgłaszaniu awarii do serwisu, przez co czas ich trwania jest dłuższy [4]. Duża liczba alarmów powoduje również negatywne skutki takie jak przeciążenie informacyjne operatorów, wynikające z tego błędy obsługi oraz wzrost kosztów bieżącej eksploatacji systemu. Wyniki analizy aktualnego stanu funkcjonowania systemu monitorującego były główną przesłanką do podjęcia badań nad problemem rozpoznania alarmów i ich klasyfikacji.

Omawiany w artykule system diagnostyczny funkcjonuje w jednym z zakładów górniczych należących do KGHM Polska Miedź S. A. Informuje on operatora o uszkodzeniach mających status usterki lub awarii. Awarią nazwiemy skutek zdarzenia powodujący przerwę w funkcjonowaniu łączności. Za usterkę uznaje się każde uszkodzenie, którego skutek jest inny niż utrata łączności [4].

Celem publikacji jest przedstawienie metody automatycznego rozpoznawania alarmów i ich klasyfikacji pozwalającej na ekstrakcję nowych i użytecznych informacji oraz reguł identyfikacji, w szczególności fałszywych alarmów generowanych przez system monitorujący.

W artykule opisano system telekomunikacyjny kopalni, którego dotyczy problem fałszywych alarmów, zastosowaną metodę ekstrakcji użytecznej wiedzy z danych oraz wyniki eksperymentów, które zostały przeprowadzone na rzeczywistych danych pochodzących z bazy systemu.

2. Sieć telekomunikacyjna w kopalni

W systemach telekomunikacyjnych funkcjonujących w kopalniach KGHM Polska Miedź S.A. zastosowano rozwiązania techniczne i organizacyjne zapewniające bardzo wysoki poziom niezawodności ich pracy. Większość systemów to bardzo nowoczesne i innowacyjne rozwiązania. Najistotniejsze z nich to: system łączności ogólnokopalnianej, system rozgłoszeniowo-alarmowy i system łączności radiowej, którego baza danych została wykorzystana do ilustracji metody rozpoznawania fałszywych alarmów.

System łączności radiowej składa się z komponentów fizycznych, jak i pakietów programowych, które w ramach jednego systemu teleinformatycznego umożliwiają realizację funkcji telekomunikacyjnych, administracyjno-serwisowych oraz diagnostycznych.

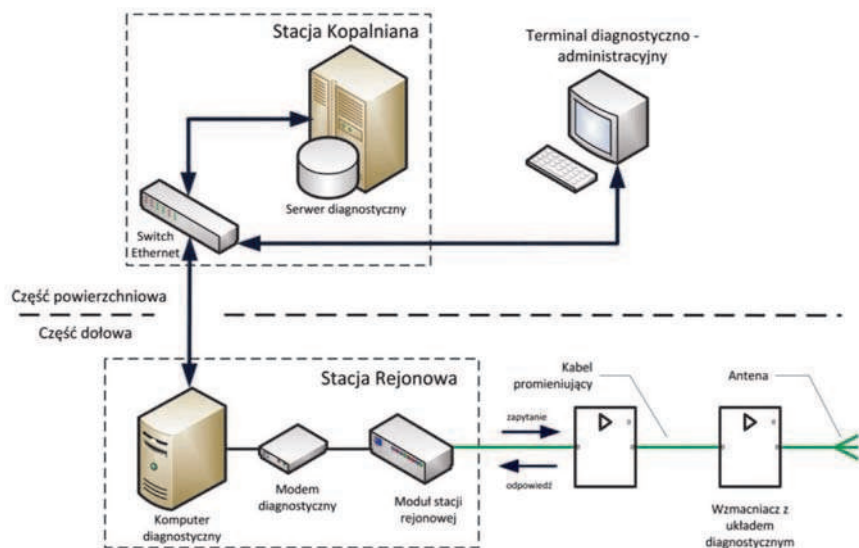
Głównym elementem systemu jest Stacja Kopalniana pełniąca rolę węzła zestawiającego połączenia, serwera aplikacji diagnostycznej oraz punktu podłączenia terminali administracyjnych. Stacja Kopalniana jest połączona łąkami światłowodowymi ze Stacjami Rejonowymi zainstalowanymi na dole kopalni. Stacje Rejonowe wyposażone są w moduły oraz modemy diagnostyczne, które „odpytują” urządzenia sieciowe w celu określenia ich statusu komunikacyjnego oraz kontroli parametrów pracy. „Zapytanie” o „obecność” danego urządzenia jest wysyłane co około 11 minut, w przypadku braku odpowiedzi aktywowany jest alarm o braku komunikacji z danym urządzeniem. Opisane elementy systemu monitorującego zilustrowano na rysunku 1.

Opisywany system łączności radiowej jest nowoczesnym rozwiązaniem pod względem sprzętowym jak i aplikacyjnym, o wysokim stopniu niezawodności dedykowanym dla kopalni głębinowych. Szczególnym walorem jest graficzny interfejs Portalu Diagnostycznego pozwalający w szybki i intuicyjny sposób lokalizować miejsca wystąpienia awarii w sieci. Jedną z interesujących funkcji systemu jest możliwość „odpytania” *ad hoc* wybranego urządzenia przez operatora systemu, co znacznie przyspiesza wykonywanie prac serwisowych. Okno mapy Portalu Diagnostycznego przedstawiono na rysunku 2. Jak można zauważyć, kolorem czerwonym sygnalizowana jest awaria elementów infrastruktury o braku komunikacji; informację tę można odczytać w oknie dialogowym wywołanym przez operatora po kliknięciu na dane urządzenie. Pomimo wskazanych zalet system monitorujący nie jest pozbawiony wad. Najistotniejszą z nich jest generowanie fałszywych alarmów o braku łączności w wyrobiskach.

Po tym krótkim wprowadzeniu, w następnym rozdziale omówimy zaproponowaną metodę eksploracji danych, uzasadnimy wybór klasyfikatora oraz przedstawimy wybraną metodykę procesu eksploracji danych.

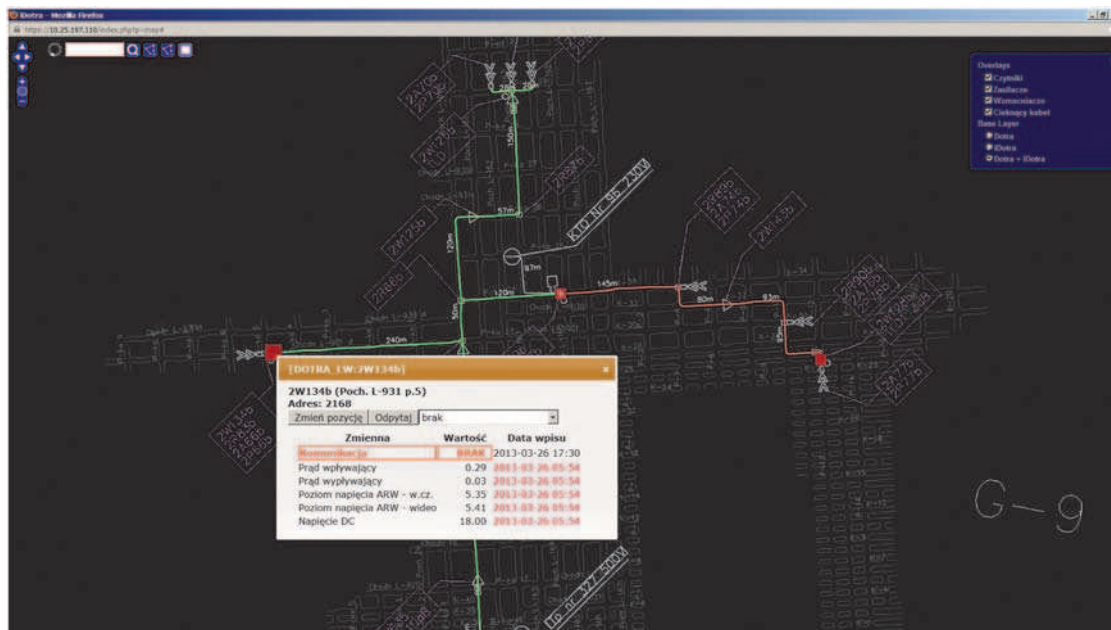
3. Metoda eksploracji baz danych diagnostycznych

Proces eksploracji danych lub odkrywania wiedzy G. Piatetsky-Shapiro definiuje jako „proces nietrywialnej ekstrakcji potencjalnie użytecznych i wcześniej nie znanych informacji bądź wyszukiwania prawidłowości i ogólnych wzorców w istniejących bazach danych” [7].



Rys. 1. Schemat poglądowy systemu monitorującego

Fig. 1. Scheme of a monitoring system



Rys. 2. Okno mapy systemu antenowego oddziału górniczego wraz z uaktywnionym oknem parametrów wzmacniacza (skala 1:7500)

Fig. 2. Place of antenna system map of mining division along with the active place of the parameters of an amplifier (scale 1:7 500)

Nawiązując do specyfiki problemu rozpoznawania i klasyfikacji alarmów, możliwe jest zastosowanie dwóch podejść dobrze opisanych w literaturze: klasyfikacji nadzorowanej lub nienadzorowanej. Zasadniczą różnicą pomiędzy nimi jest wiedza *a priori* o wartości zmiennej celu. W przypadku algorytmów nienadzorowanych taka zmienna może być utworzona po grupowaniu, a jej klasa będzie etykietą danej grupy obserwacji. Z uwagi na posiadanie informacji czy dany alarm historyczny był prawdziwy lub fałszywy, wskazane było zastosowanie klasyfikacji nadzorowanej.

Istnieje wiele metod klasyfikacji nadzorowanej, do których należą m.in.: regresja logistyczna, analiza dyskryminacyjna, sieci neuronowe, algorytmy genetyczne, SVM, naiwny klasyfikator Bayesa, reguły decyzyjne CN2, indukcyjne drzewa decyzyjne. Niemożliwe jest zastosowanie wielu z tych metod, które są dedykowane dla modeli z ilościowymi zmiennymi objaśniającymi, podczas gdy alarm opisany jest w większości za pomocą atrybutów symbolicznych. Natomiast metody takie jak SVM, naiwny klasyfikator Bayesa, reguły decyzyjne i indukcyjne drzewa decyzyjne nie mają tego ograniczenia, dlatego ta grupa metod była rozpatrywana w celu wskazania tej, która ze względów teoretycznych, jak i praktycznych jest odpowiednia dla uzyskania oczekiwanego rezultatu.

Przeprowadziliśmy eksperymenty i testy wielu metod klasyfikacji, wśród których wyróżniliśmy indukcyjne drzewa decyzyjne. Za wyborem tej metody przemawiało kilka przesłanek. Pierwsza wynikała z podstawowej zasady wnioskowania indukcyjnego prowadzącej do generalizacji obserwacji i faktów w postaci reguł i stwierdzeń. Analitik posiadający wiedzę dziedzinową powinien weryfikować prawdziwość wygenerowanych reguł i modelować drzewo dopóty, dopóki reguły te będą na tyle pragmatyczne, by zastosować je do rozwiązania problemu. Następnym ważnym walorem była prostota interpretacyjna otrzymanych reguł, zarówno w postaci graficznej, jak i reguł decyzyjnych. Ostatnią zaletą, która skłoniła nas ku zastosowaniu drzew indukcyjnych jest możliwość kontroli złożoności i ogólności generowanych reguł. Słabością IDT jest możliwość wygenerowania zbyt „rozrośniętego” drzewa,

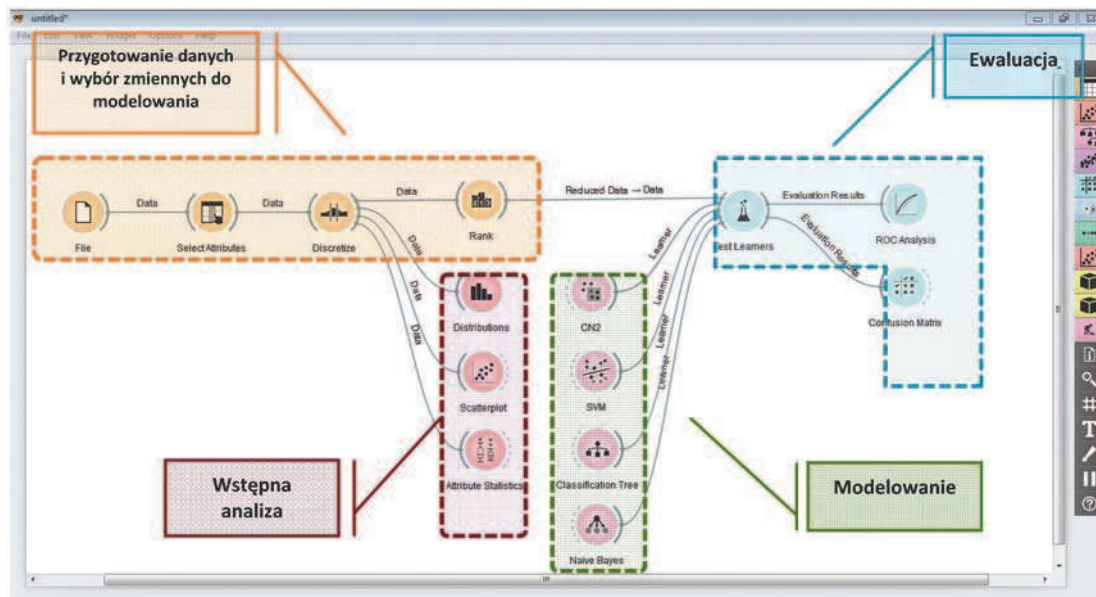
które może być przeuczone (*overfitting*) i bardziej podatne na błędne klasyfikacje nowych przypadków [1].

Istnieje wiele miar oceny klasyfikatorów takich jak: wrażliwość (*sensitivity*, *TP rate*), precyzja (*precision*), specyficzność (*specificity*, *TN rate*) lub trafność (*accuracy*) [6]. W przeprowadzonych badaniach w aspekcie wartości wskazanych miar najistotniejszym celem była minimalizacja błędów pierwszego rodzaju (FP), określającego liczbę fałszywych alarmów, które zostały zaklasyfikowane jako prawdziwe¹. Drugą istotną miarą była precyzja, która uwzględnia liczbę alarmów prawdziwych błędnie zaklasyfikowanych jako fałszywe. Generalnie przy wyborze klasyfikatora dążymy do osiągnięcia kompromisu pomiędzy czytelnością oraz użytecznością reguł i maksymalizacją wartości tych miar, przy czym pierwsze kryterium należy uznać za bardziej istotne w przypadku niewielkiej różnicy w ocenie klasyfikacji.

Proces eksploracji danych został przeprowadzony według metodyki CRISP-DM [1], przy użyciu platformy Orange. W badaniu zaproponowano schemat procesu eksploracji przedstawiony na rysunku 3. Pierwszym etapem było przygotowanie danych do modelowania, którego część przeprowadzono w MS EXCEL i MS ACCES. Tę część procesu wyróżniono na omawianym rysunku pomarańczowym kolorem. Kolejnym etapem procesu było przeprowadzenie wstępnej analizy danych (pole z brązowym konturem), po którym zbudowano model i przeprowadzono właściwą eksplorację danych za pomocą wybranych metod klasyfikacji (pole z zielonym konturem). Ostatnim etapem była ewaluacja wybranych modeli (pole z niebieskim konturem).

Opis danych, proces budowy modelu indukcyjnego drzewa decyzyjnego i jego oceny w porównaniu z innym klasyfikatorami (SVM, Naive Bayes i CN2) zostaną przedstawione w następnym rozdziale.

¹ W tabelarycznym zestawieniu parametrów oceny jakości klasyfikatorów zaprezentowanym w kolejnym rozdziale posłużyliśmy się parametrem wrażliwości (1-FP), określa on liczbę poprawnie zaklasyfikowanych fałszywych alarmów.



Rys. 3. Okno procesu eksploracji danych platformy Orange
Fig. 3. Place of data exploration on the Orange platform

4. Wyniki eksperymentów

Analizowane dane są danymi rzeczywistymi pochodzącymi z: systemu monitorującego, dokumentacji technicznej systemu oraz wpisów w książce serwisowej. Zbiór danych stanowił 1316 obserwacji z okresu od 01.05.2012 r. do 18.05.2012 r. Opisywana próbka danych została wybrana losowo. W pierwszym etapie prac w celu wykorzystania danych w wybranych algorytmach klasyfikacji przeprowadzono niezbędne transformacje wartości atrybutów. Przykładowo, jeden z atrybutów („Czas trwania”) poddano dyskretyzacji (metodą równej gęstości klas). Przekształcone dane poddano następnie rangowaniu, korzystając z kryterium Gain Ratio, następnie wybrano atrybuty, które zostały użyte do modelowania.

Jak wskazano i uzasadniono w rozdziale trzecim, do modelowania klasyfikatora wybrano algorytm indukcyjnego drzewa decyzyjnego opartego na entropii (ID3) [3]. W trakcie eksperymentów wygenerowano około 100 wariantów drzew decyzyjnych, zmieniając wartość różnych parametrów wstępnego przycinania drzewa oraz liczbę klas atrybutu opisującego czas trwania alarmu.

W pierwszej serii eksperymentów określono liczbę klas atrybutu „Czas trwania” na 10 (maksymalną dla ograniczenia w pakiecie Orange). Parametrem kontrolowanym był warunek wstępnego przycinania drzewa (minimalna liczba obserwacji w liściu), którego wartość po przeprowadzeniu 10 eksperymentów ustalono na 5, zakładając, że jego wartość może być ponownie zmieniana podczas ewaluacji modelu.

Na skutek wygenerowania drzewa, w którym 6 liści węzła atrybutu „Czas trwania” zawierało tylko obserwacje z fałszywymi alarmami, dokonano agregacji tych klas, czego efektem było zmniejszenie ich liczby z 10 do 5. Wskazana

zmiana negatywnie wpłynęła na jakość wszystkich czterech klasyfikatorów. Parametr specyficzności (TN) przyjął gorszą wartość niż w poprzednich eksperymentach.

W ostatniej serii eksperymentów ponownie dokonano zmiany parametru zatrzymania budowy drzewa. Tak samo jak w pierwszej serii eksperymentów podczas badania kontrolowano liczbę obserwacji w liściu, przy której budowa drzewa jest zatrzymywana. Parametr ten przyjmował wartości w przedziale od 1 do 10. Satisfakcjonujące rezultaty w kontekście wartości miar oceny i użyteczności wygenerowanych reguł uzyskano dla wartości warunku zatrzymania budowy drzewa równej dwóm obserwacjom.

W tabelicy nr 1 przedstawiono zestawienie parametrów oceny klasyfikatorów. Jak można zauważyć, niewiele lepsze wyniki wskaźnika wrażliwości zostały uzyskane dla klasyfikatorów SVM i CN2, o wiele gorszy wynik osiągnięto dla Naive Bayes. Wartość precyzji klasyfikacji dla indukcyjnych drzew decyzyjnych była najlepsza wśród zaproponowanych klasyfikatorów. Specyficzność określająca poprawność klasyfikacji prawdziwych alarmów była niewiele gorsza niż w przypadku SVM, jednocześnie lepsza niż dla CN2 i Naive Bayes. Ostatnią, najmniej istotną miarą, ze względu na różne znaczenie błędnych klasyfikacji dla każdego rodzaju alarmu była trafność klasyfikacji. W przypadku tej miary wynik był nieznacznie gorszy niż dla CN2 i SVM. Należy podkreślić, że parametry oceny klasyfikacji były bardzo zbliżone dla drzew decyzyjnych, CN2 i SVM. Zdecydowanie najgorszym klasyfikatorem był Naive Bayes. Uwzględniając kryteria miar oceny jakości klasyfikacji, łatwości interpretacyjnej reguł i użyteczności nowej wiedzy na temat fałszywych alarmów, metoda indukcyjnych drzew decyzyjnych została uznana za lepszą od pozostałych, które zostały wstępnie zaproponowane.

Tablica 1. Zestawienie parametrów oceny jakości stosowanych klasyfikatorów

Table 1. Summary of parameters of the evaluation of quality of the classifiers in application

| Klasyfikator | Wrażliwość | Precyzja | Specyficzność | Trafność |
|---------------------|------------|----------|---------------|----------|
| Classification Tree | 0,972 | 0,928 | 0,555 | 0,912 |
| CN2 | 0,974 | 0,927 | 0,550 | 0,913 |
| SVM | 0,973 | 0,927 | 0,565 | 0,914 |
| Naive Bayes | 0,923 | 0,886 | 0,298 | 0,832 |

W artykule przedstawiono jedynie cztery spośród kilkunastu reguł, które dostarczyły użytecznej, nieznanej wcześniej wiedzy na temat fałszywych oraz prawdziwych alarmów generowanych przez system monitorujący. W około 100 eksperymentach wygenerowano kilkaset reguł decyzyjnych².

Analizując wygenerowane reguły odnotowano, że w każdym z drzew wygenerowanych po agregacji klas „Czas trwania alarmu” (trzecia seria eksperymentów) występuje następująca „silna” reguła o „czystości” 96,5% pokrywająca 800 alarmów:

JEŻELI

„Moment wystąpienia alarmu: praca kopalni” = „Wydobycie”

\wedge „Czas trwania alarmu” \leq 44 minuty

TO Alarm fałszywy

Wskazana reguła jest nową użyteczną wiedzą na temat progu czasowego alarmu, poniżej którego alarm można z wysokim prawdopodobieństwem uznać za fałszywy. Wartość tego progu wynosi około 44 minut.

Kolejna użyteczna reguła, pomimo pokrycia znacznie mniejszej liczby przypadków to:

JEŻELI

„Moment wystąpienia alarmu: praca kopalni” = „Postój”

\wedge „Moment ustąpienia alarmu: praca ludzi” = („Zakładka WSP” „Zmiana wydobywczą”)

TO Alarm prawdziwy

Wskazana reguła pokrywa 39 przypadków, z „czystością” 100%. Po przeanalizowaniu przypadków, których dotyczy, okazało się, że identyfikuje ona alarmy, które powstały na skutek wyłączenia rozdzielni elektrycznych w weekend. Zastosowanie tej reguły pozwoliło na zwiększenie skuteczności rozpoznawania alarmów dla osoby prowadzącej rozliczenia wskazać rozdzielnie, które ze względu na zasilanie urządzeń łączności powinny być utrzymywane w stanie ciągłej pracy. Informacja ta powinna być przekazana służbom elektrycznym w kopalni.

Kolejne dwie reguły klasyfikują alarmy o długim czasie trwania, należy zaznaczyć, że w praktyce szczególnie trudno określić prawdziwość takich alarmów.

JEŻELI

„Moment wystąpienia alarmu: praca kopalni” = „Wydobycie”

\wedge „Czas trwania alarmu” $>$ 5 godzin i 30 minut

\wedge „Moment wystąpienia alarmu: praca ludzi” = „Zmiana wydobywczą”

\wedge „Moment ustąpienia alarmu: praca ludzi” = „Brak wydobywania”

TO Alarm fałszywy

JEŻELI

„Moment wystąpienia alarmu: praca kopalni” = „Wydobycie”

\wedge „Czas trwania alarmu” $>$ 5 godzin i 30 minut

\wedge „Moment wystąpienia alarmu: praca ludzi” = („Wydobycie” „Prace strażowe”)

TO Alarm prawdziwy

Wskazane reguły dostarczają użytecznej wiedzy na temat fałszywych alarmów o długim czasie trwania. Jak wskazano we wstępie, operatorzy często posługują się heurystykami przy próbie określenia prawdziwości alarmu, dotychczas zakładano, że alarm fałszywy nie trwa dłużej niż 2 godziny. Na podstawie pierwszej z reguł można wywnioskować, że w 45 przypadkach alarmy o czasie trwania około 5,5 godziny były alarmami fałszywymi („czystość” liścia = 100 %), co

podważa dotychczasowe, heurystyczne założenia. Druga z reguł pozwala identyfikować przypadki, gdy do utraty łączności dochodzi najprawdopodobniej na skutek prowadzenia prac strażowych. Reguła ta pokrywa tylko trzy przypadki („czystość” liścia 100%), jednak biorąc pod uwagę informacje jakich dostarcza, może identyfikować sytuacje, w których dochodzi do uszkodzenia linii telekomunikacyjnej na skutek prowadzenia prac strażowych. Takie informacje przekazane służbom strażowym mogą zmniejszyć częstość występowania takich sytuacji, co pozwoli na zmniejszenie awaryjności systemu oraz zmniejszy koszty związane z serwisem uszkodzeń sieci.

5. Podsumowanie

Systemy telekomunikacyjne w górnictwie pełnią kluczową rolę w aspekcie bezpieczeństwa, dobrej organizacji i utrzymania ciągłości wydobywania, dlatego konieczne jest monitorowanie stanu infrastruktury telekomunikacyjnej. Pomimo nowoczesnych i innowacyjnych rozwiązań omawiany system monitorujący nie jest wolny od wad. W artykule zaproponowano metodę wykrywania fałszywych alarmów o braku łączności w wyrobiskach górniczych oraz poszukiwania nowej, użytecznej wiedzy z danych. W około 100 eksperymentach przy kontrolowaniu warunków wstępnego zatrzymania budowy drzewa oraz liczby klas atrybutu „Czas trwania alarmu” za pomocą indukcyjnych drzew decyzyjnych pozyskano informację o wartości progowej czasu trwania, poniżej której alarm obiektywnie można uznać za fałszywy. Kolejną bardzo istotną, nieznaną wcześniej informacją jest zidentyfikowanie pojedynczych urządzeń, które zostają pozbawione napięcia zasilania w dni, w które nie jest prowadzone wydobywanie, co powoduje utratę łączności. Inna potencjalnie użyteczna reguła, dotyczy rozpoznawania urządzeń, które mogą być uszkodzone na skutek prowadzenia prac strażowych.

Otrzymane wyniki skłaniają do podjęcia dalszych badań nad rozpoznaniem i klasyfikacją alarmów sygnalizowanych przez system monitorujący w kopalni. Przede wszystkim badania należy przeprowadzić na znacznie większej liczbie obserwacji. W celu poprawienia trafności klasyfikacji należy rozważyć dodanie kolejnych atrybutów opisujących alarm oraz uwzględnienia w bazie danych systemu diagnostycznego połączeń pomiędzy wzmacniaczami oraz kolejności wystąpienia stanów awaryjnych w czasie. Efekty tych prac pozwolą na sprawniejsze i bardziej efektywne funkcjonowanie systemu monitorowania stanu instalacji telekomunikacyjnej.

Literatura

1. *Bramer M.*: Undergraduate Topics in Computer Science. Principles of Data Mining. Springer. London 2013, s. 121÷122.
2. *Ding S.*: Model-Based Fault Diagnosis Techniques. Design Schemes, Algorithms and Tools, Springer. London 2013, s. 4.
3. *Gorunescu F.*: Data Mining. Concepts, Models and Techniques. Springer. Berlin 2011, s. 165÷166.
4. *Karaban B.*: Indukcyjne drzewa decyzyjne w analizie alarmów systemu telekomunikacyjnego. Praca magisterska. UE we Wrocławiu. Wrocław 2013.
5. *Korbicz J., Kościelny J., Kowalczyk Z., Cholewa W.*: Diagnostyka procesów. Modele. Metody sztucznej inteligencji. Zastosowania. WNT. Warszawa 2002, s. 11.
6. *Morzy T.*: Eksploracja danych. Metody i algorytmy. PWN. Warszawa 2013, s. 326÷327.
7. *Piatetsky-Shapiro G., Frawley W.*: Knowledge discovery in databases. The AAAI Press. Menlo Park 1991.

² Kompletnie wyniki eksperymentów przedstawiono w pracy magisterskiej B. Karabana „Indukcyjne drzewa decyzyjne w analizie alarmów systemu telekomunikacyjnego”.